NBER WORKING PAPER SERIES

MARGINALIZED AND OVERLOOKED? MINORITIZED GROUPS AND THE ADOPTION OF NEW SCIENTIFIC IDEAS

Wei Cheng
Bruce A. Weinberg

Working Paper 29179
http://www.nber.org/papers/w29179

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2021

We thank seminar participants at Ohio State University. We gratefully acknowledge support from NIA, OBSSR, and NSF SciSIP through P01 AG039347; R01 GM140281, UL1 TR002733, NSF EHR DGE 1348691, 1535399, 1760544, 2100234. Weinberg was supported on P01 AG039347 by the NBER directly and on a subaward from NBER to Ohio State. Cheng was supported by National Science Foundation of China (71803047) and Shanghai Pujiang Program (2019PJC022). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Marginalized and Overlooked? Minoritized Groups and the Adoption of New Scientific Ideas
Wei Cheng and Bruce A. Weinberg
NBER Working Paper No. 29179
August 2021
JEL No. D85,J71,O31

ABSTRACT

The rapid diffusion and use of new ideas are critical for advancing and realizing the value of innovation. This paper explores the impact of demographic characteristics of innovators and potential adopters on the adoption of important new scientific ideas through networks. Using rich, population-level data on the biomedical researchers and their networks, natural language processing, and a novel two-way fixed effects strategy, we find that new ideas introduced by female scientists are under-utilized, which can be explain by two factors. First, female innovators are not as well-connected in networks; second, even at a short network distances, researchers (especially men) are less likely to adopt women's ideas. Ideas from underrepresented racial and ethnic minorities are also less widely used.

Wei Cheng
School of Business
East China University of Science
and Technology
130 Meilong Rd, XuHui
Shanghai, China 200037
chengwei@ecust.edu.cn

Bruce A. Weinberg
The Ohio State University
Department of Economics
410 Arps Hall
1945 North High Street
Columbus, OH 43210
and NBER
weinberg.27@osu.edu

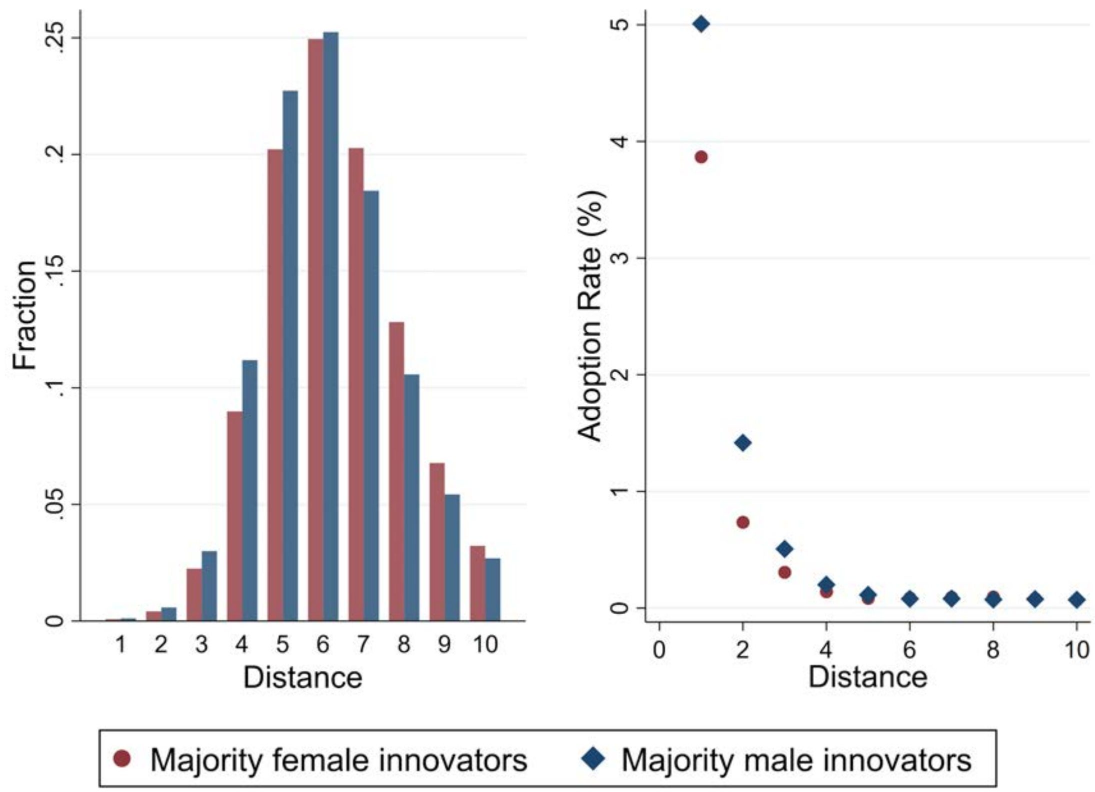
1. Introduction

Do women and underrepresented racial and ethnic minorities receive less attention than men and majority groups for their contributions (Wenneras and Wold, 1997; Halpern, et al., 2007; Koffi, 2021)? This paper explores this question in an interesting and consequential population - biomedical scientists. Scientists are ideal for studying flows of ideas because they are highly knowledgeable and strongly incentivized to keep abreast of and adopt the latest advances. Moreover, the contents of scientists' contributions are publicly available, their contributions can be identified from authorship, and their networks can be reconstructed. With the population of scientists becoming increasingly diverse, disparities in the adoption of ideas are becoming increasingly costly for the scientific enterprise. Biomedical scientists are particularly consequential because research on conditions relevant for underrepresented groups is frequently undertaken by members of those groups, so a failure of researchers from these groups to receive appropriate attention would likely increase underlying health disparities (Oh, et al., 2015; Association of American Medical Colleges, 2009).

This paper combines two methods to explore differences by gender, race, ethnicity, and age in the adoption of ideas. First, we employ recent methods in natural language processing to identify the most important ideas (one-, two-, and three-word combinations, "n-grams") introduced each year between 1980 and 2008 and track their utilization at population scale for nearly a quarter of a century (see Packalen and Bhattacharya, 2019; Staudt et al., 2018). We do so using the text of titles and abstracts in MEDLINE, a massive database of biomedical research articles with nearly 30 million articles dating back to the 19th century. The ideas we study are important. Among many other important advances, they include the work on HIV / AIDS that has transformed it from an incurable condition to a chronic, manageable disease. They also cover the development of a wide range of methods and findings in molecular biology that are the foundation for the biotech revolution. Second, we build on the well-known idea that technology adoption declines with distance. Using these building blocks, we identify as “innovators” all of the authors of article(s) that first introduce an idea that becomes important. We then study how the adoption of ideas decays with network distance differentially for innovators and potential adopters but demographic characteristics.

Figure 1 illustrates our approach in the case of gender, which has implications for women’s health. The left panel shows that researchers are on average further from ideas generated by teams with more women than men (red bars) compared to ideas generated by teams with more men than women (blue bars). This finding implies that female innovators are not as well-connected as male innovators. What further exacerbates disparities is that at each distance, ideas introduced by female majority teams are less likely to be adopted (right panel). Even at a distance of 1, people are more likely to choose to build on ideas introduced by majority male teams. Both effects lead to lower adoption rates for female majority new ideas. Specifically, the 5-year adoption rate of new ideas from female majority teams is 23% lower than that of ideas from male majority teams (even among ideas in the top .1% of ideas). Of this difference, 32% is due to less proximity to the women-majority inventor teams, but the considerable majority (68%) is due to lower utilization at any given distance.

Figure 1. Gender composition of innovator team



Notes: All new ideas are divided into two groups: ideas introduced by teams with a majority of female innovators (red) and ideas introduced by teams with a majority of male innovators (blue). The left panel depicts the distribution of researchers' network distance to the two groups of ideas separately. The right panel shows the 5-year adoption rates for ideas from both sets of teams in percentage points at each network distance.

Although we have no reason to believe it is the case, one potential explanation for our finding that the ideas introduced by women are adopted less than those introduced by men is that the ideas introduced by women may be inherently different (e.g. not be as promising or widely applicable) as those introduced by men. To address this possibility, we rely on two key features of our context: (1) that there is typically more than one author on an article and (2) that network proximity is a strong determinant of knowledge flows. In our estimation, for each idea and each potential adopter, we identify the nearest individual innovator. We estimate regressions that control for idea fixed effects that eliminate all differences across ideas in the applicability of the idea to potential adopters. Intuitively, each active biomedical researcher (i.e. potential adopter) will be closer to one (or more) of the innovators than to others. Thus, we can leverage the network structure, measured prior to the introduction of the idea, to identify whether, for each idea and potential adopter, the "closest" innovator to that potential adopter is a woman or a man, and see whether, controlling for the distance to the closest innovator, each potential adopter is more likely to adopt an idea when the innovator who is closest to them is a woman or a man. In this way, we use the presence of multiple innovators and network proximity to control for the "quality" of ideas.

Another potential explanation for our finding that the people who are closer to women innovators are less likely to adopt ideas is that the people closer to women innovators are inherently less likely to adopt new ideas. For instance, women tend to be closer to other women, and women may be less likely to adopt new ideas, perhaps because they occupy positions that are less central in networks (and hence have less access to complementary innovations) or because of less access to resources (Ginther and Kahn, 2009).[1] We address these concerns by directly measuring network distances. We also observe adoption behavior for multiple ideas by each potential adopter and include fixed effects for the adoption probability of each potential adopter. In this way, we control for unobserved differences in adoption propensities for each potential adopter. Our estimates indicate that, at any given distance, women and underrepresented minorities (URMs) are less likely to adopt new ideas. However, controlling for these differences and potential adopter fixed effects accounts for only a small portion of the disparity in the adoption of ideas of women innovators.

[1] Another potential concern is that our sample of ideas is itself determined by adoption rates, so that if the characteristics of innovators do indeed determine adoption, the characteristics of innovators may well affect inclusion in our sample. We acknowledge that such a process operates, but note that we (1) control for idea fixed effects, so that our estimates are identified from within-idea variations and (2) that this process implies that the ideas from, for instance, predominantly women innovator teams in our sample are, if anything, "better" ideas than those from predominantly men innovator teams.

Our work contributes to the literature in three ways. First, we contribute to the literature on underrepresentation in science. While the pool of scientific workers is increasingly diverse, women and underrepresented racial and ethnic groups are often marginalized, and the health conditions specific to them are understudied (Oh, et al., 2015). [2] A sizeable body of work on underrepresentation focuses on disparities across scientists from different groups in terms of access to resources and compensation (e.g. Ginther and Kahn, 2009; Buffington, et al., 2016). Increasingly, researchers are focusing on the credit that women receive for their contributions in the form of author order (e.g. Cikara et al., 2012, West, et al. 2013, Lerchenmüller, Lerchenmüller, and Sorenson, 2018, Marschke, et al., 2018), in terms of inclusion as an author at all (Ross et al. 2021), or credit for authorship (e.g., Sarsons et al. 2021). We explore another approach - the possibility that the ideas produced by women and underrepresented groups simply receive less attention in the literature because they are produced by women and/or URMs. Perhaps the closest to our analysis is a concurrent analysis by Koffi (2021), which focuses on citations in economics.

Second, our research also connects to the literature on knowledge spillovers. Technology and knowledge spillovers have been central to economists' understanding of growth, trade, and urban agglomerations for over thirty years (Romer, 1986; Lucas, 1988; Krugman, 1991; and Glaeser, et al., 1992). Unfortunately, the micro evidence for such spillovers tends to be indirect (e.g. inferred from productivity) rather than measured directly and is surprisingly weak given the critical importance of knowledge spillovers (Azoulay, Graff-Zivin and Wang, 2010; Waldinger, 2010, 2012; Borjas and Doran, 2012, 2015; and Catalini, 2017).[3] One novel aspect of our work relative

[2] Somewhat more distantly related to our work, it is generally agreed that intellectual diversity contributes to productivity and creativity by increasing the breadth of ideas (Freeman and Huang, 2014) and providing a broader range of potential solutions to a problem (Rhoten et al., 2009). While intellectual diversity and demographic diversity are conceptually distinct, they are positively associated in practice (Reagans and Zuckerman, 2001).

[3] Kaiser (2005) is a fascinating exception, showing how the use of Feynman diagrams diffused through the physics community from the different ways in which physicists drew them.

to this literature is that we directly measure knowledge flows using text analysis. At least since the seminal work by Jaffe, Trajtenberg and Henderson (1993), researchers have sought to infer knowledge flows using citations, frequently in patent records across countries / regions (Peri, 2005), between firms (Almeida and Kogut, 1999; Gomes-Casseres, 2006), and among scientists (Singh, 2005). Text analysis has a number of advantages over this approach - it allows us to track the use of specific ideas and avoid artifacts in the citation process in both patents[4] and the scientific literature[5].

Lastly, our analysis of coauthorship networks allows us to focus on transmission of ideas across individuals who have connections with one another. Given that people's knowledge of new ideas is imperfect, researchers close to the innovators of a new idea benefit from the ability to obtain tacit knowledge at a lower cost. The literature offers abundant evidence regarding the importance of *spatial proximity* in knowledge flow (e.g. Jaffe, et al., 1993; Almeida and Kogut, 1999; Thompson and Fox-Kean, 2005, Waldinger, 2010, 2012; Borjas and Doran, 2012, 2015; Catalini, 2017; Ham and Weinberg, 2021). However, most of the mechanisms underlying the role for spatial proximity, including a higher chance of meeting, lower communication costs, and enhanced trust, hold for *network proximity* as well. Despite its potential as a mediator, the literature on network proximity for idea diffusion is relatively limited (Singh, 2005; Zacchia, 2019) and our two-way fixed effects design has the potential for facilitating future network analyses.

[4] Alcácer and Gittelman (2006) show that more than half of the citations on a patent are added by patent examiners, and thus may not represent knowledge flows.

[5] Following Merton (1973), scientific citations are taken to reflect knowledge flows, but there is considerable evidence for a variety of other motives for citations (see Gilbert, 1977; Moed, 2005; Moed and Garfield, 2004; Collins, 2004; Woolgar, 1991; Fong and Wlhite, 2017; Koffi, 2021). Moreover, different types of articles receive different numbers of citations (Moed and van Leeuwen, 1995).

2. Data

2.1 Data

Our primary data source is MEDLINE, the U.S. National Library of Medline’s (NLM) premier bibliographic database, which is akin to Econlit for biomedical research. MEDLINE contains nearly 30 million biomedical research articles from around the world dating back to the 19th century, with coverage that improves greatly over time, especially after 1966. In addition to the comprehensiveness of the MEDLINE data, another strength is that it provides rich information on journal articles, including authors’ names, publication year, and the text of titles and abstracts, which we use in our text analysis. MEDLINE also includes Medical Subject Headings (MeSH) for each publication that indicate its main research areas. We use these to identify potential adopters of a new idea by identifying researchers who have some work in common with the innovators who first introduced the idea.

Another crucial data source is the “Author-ity” MEDLINE author disambiguation database, updated through 2018 (Torvik, Weeber, Swanson, and Smalheiser, 2006; Torvik and Smalheiser, 2009). In a corpus the size of MEDLINE, identifying authors solely based on names is likely to generate lumping errors, where people with similar names are treated as one person, and splitting errors, where one author is regarded as different researchers when names are written in different ways. “Author-ity” provides a high-quality, algorithmic disambiguation of MEDLINE that allows us to identify each scientist’s publication records with a high level of accuracy.

In this paper, we focus on U.S.-based scientists. In order to identify author locations, we use MapAffil (Torvik, 2015), which offers affiliation information for MEDLINE authors. We eliminate people who are ever affiliated with organizations outside the U.S. Additionally, we supplement our data with Genni (Torvik and Agarwal, 2016), which predicts gender, and Ethnicolr (Laohaprapanon and Sood, 2017), which predicts race and ethnicity based on scientists' names. Genni provides ethnicity-specific gender predictions. Ethnicolr classifies people into four categories that combine race and ethnicity: Hispanics (of any race), non-Hispanic Asians, non-Hispanic Blacks, and non-Hispanic Whites.

2.2 List of new ideas

Following Packalen and Bhattacharya (2019) and Staudt et al. (2018), we extract the main concepts or ideas contained in a paper by taking the words (unigrams), two-word phrases (bigrams) and word triplets (trigrams) in the text of the title and abstract. A new idea is an n-gram that first appeared in a certain year and that had never been used before. We refer to this year of first appearance as the birth year of a new idea. We focus on new ideas born between 1980 and 2008. This time window is chosen for two reasons. First, we exclude ideas that are born before 1980 to allow for a sufficient "burn-in" period in the calculation of an idea's birth year. Many old terms will be identified as having a birth year when the data start[6]. However, after over ten years, a first-appearing term is likely to be an innovation. Second, the last idea birth year we choose is 2008 so that up to 2018 there are at least ten years after an idea is born to observe its adoption.

One potential problem with this approach of using new terms to represent new ideas is that some new terms are merely new combinations of words but do not constitute important innovations. To avoid such cases, which are rarely widely adopted, we rank the ideas by the number of total

[6] In the early years of the data, terms such as "blood" have very high frequencies of mentions. However, they are not innovations but commonly used language in academia. Hence, it is important to have a sufficiently long initial period to set them apart from new ideas.

mentions within each birth year and take the top 0.1% of the distribution for each idea birth year. Therefore, the final list of new ideas we focus on is a set of the most important new ideas in biomedicine originated by U.S. scientists. For example, ideas ranked highly include well-known breakthroughs and advances in biomedical research such as polymerase chain reaction, small interfering RNAs, and HIV / AIDS. The number of the important new ideas in our list ranges from 105 in 1980 to 281 in 2008.

2.3 Sample construction

For each new idea (and its innovators), we identify potential adopters by selecting all other researchers who have works with at least one MeSH term in common with the innovators before the birth of the new idea. Formally, to construct our sample for idea i that first appears (i.e. is born) in year τ, we let $\mathbb{I}_{i,\tau}$ denote the set of innovators who originated idea i in year τ. Let $M_{i,\tau}$represent all MeSH terms that $\mathbb{I}_{i,\tau}$ have published on in or before year τ. Our set of potential adopters for idea i, $\mathbb{A}_{i,\tau}$, is defined as the set of researchers who have published on at least one MeSH term in $M_{i,\tau}$in or before year τ and who are not in $\mathbb{I}_{i,\tau}$ itself.

Next, for each potential adopter in $\mathbb{A}_{i,\tau}$, we calculate his/her shortest path to any of the innovators of idea i, $\mathbb{I}_{i,\tau}$, based on the collaboration networks in the year the idea was introduced and the two preceding years, $\tau - 2$ to year τ[7]. (Adoption is tracked starting the year after τ, i.e. from $\tau + 1$.) Network distance represents how close a potential adopter is to a new idea in collaboration networks. (We truncate the network distance at 10 because over 90% of all finite distance pairs have a distance equal to or less than 10 and adoption rates at large distances are extremely low.)

[7] Since network links are observable only when publications are generated, and it is common for a scientist to have zero publications in some years, we choose a three-year window around the birth of a new idea to define the networks. Infinite network distance is not considered in this analysis.

Two time frames are used to measure adoption: 5 years (i.e. starting from $\tau+1$ through τ+5) and 10 years (i.e. starting from $\tau+1$ through τ+10) after introduction of a new idea. The primary results presented in the paper are based on adoption within 5 years because the effects of pre-existing network proximity likely decline as time passes. Analyses of adoption within 10 years are conducted as robustness checks. The adoption of new ideas is our primary outcome. We also study "independent adoption," which we define as using a new idea in an article that does not have one of the original innovators as a coauthor because "adoption" through coauthorship with one of the innovators of the idea likely operates differently from independent adoption.

Our unit of analysis is an idea-adopter pair, and our sample construction procedure yields over 74 million observations. We take a 10% random sample as our analytical dataset. Our final analytical sample, which is described in Table 1, consists of 7,417,333 observations, corresponding to 3,430 new idea terms and 536,987 unique potential adopters. Of all potential adopters, 52% are men, 32% are women, and the rest, 15%, have uncertain gender predictions. In terms of race and ethnicity, the largest group is non-Hispanic White (82.8%), followed by non-Hispanic Asian (12.6%), Hispanic (4.0%), and non-Hispanic Black (0.6%). These demographic features do not represent the overall scientific workforce in biomedicine in that our sample only includes people who are at a distance less than or equal to 10 from the innovators of the new important ideas and weights people who are in areas with more important new ideas more heavily than others.

3. Results

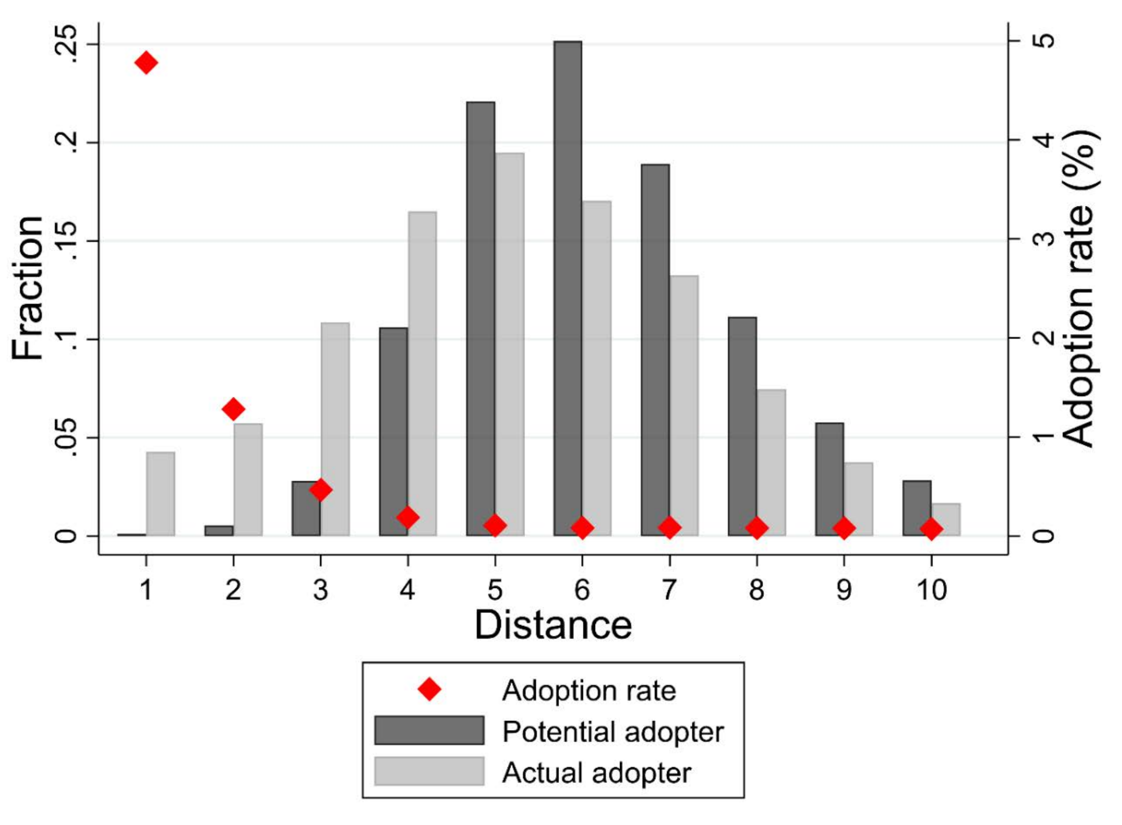
We present our results in three steps. First, we provide some graphical evidence of different patterns of idea adoption for researchers from various demographic groups. Second, we conduct a formal, regression analysis. Lastly, we perform a series of decompositions to assess the relative contributions of network distance, the gender mix of adopters, and adoption conditional on distance, to the gender gap in adoption.

3.1 Descriptive results on gender

Figure 2 illustrates the relationship between network distance and the adoption of ideas. The red squares represent the probability of adoption within 5 years of introduction (right axis) at various distances from the people who introduced the new ideas (innovators). Not surprisingly, the adoption rate of important new terms declines sharply with distance[8]. The bars and the left axis in Figure 2 show the distribution of network distance for all potential adopters (black bars) and actual adopters (grey bars). The grey bars are shifted left-ward, indicating that actual adopters are generally more proximate to innovators in network terms.

[8] It is noteworthy that while the adoption rate falls with distance, the adoption frequency is greatest at moderate distances because the number of potential adopters at moderate distances are order(s) of magnitudes higher than those at distance 1 or 2 from innovators.

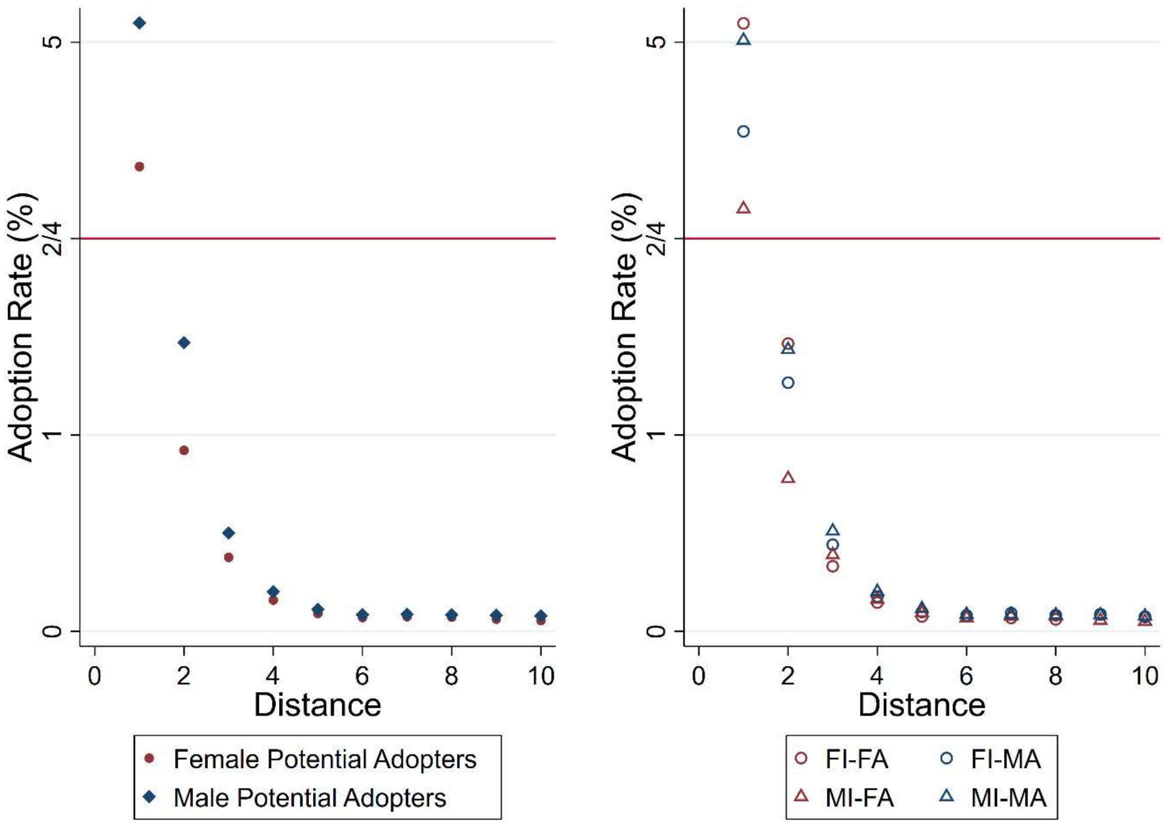
Figure 2. Adoption rates and network distance



Notes: The red diamond shapes (corresponding to the right y-axis) represent the probability of adoption within 5 years following their introduction by potential adopters at various network distances. The bars show the distribution of a potential adopter's network distance to the nearest innovator on each important new idea. The black bars represent all potential adopters and the grey bars represent actual adopters.

One reason women's ideas may be less likely to be adopted than men's ideas is that women may be more likely to adopt other women's ideas and women may be less well integrated into networks. We explore this possibility. The left panel of Figure 3 shows that women are indeed less likely to adopt important new ideas, especially at the shortest distances. On the other hand, our data shows that women and men are similarly close to new ideas in networks. The mean distance to new ideas is 6.15 for women and 6.12 for men (both have a median distance of 6).

Figure 3. Gender and adoption rates



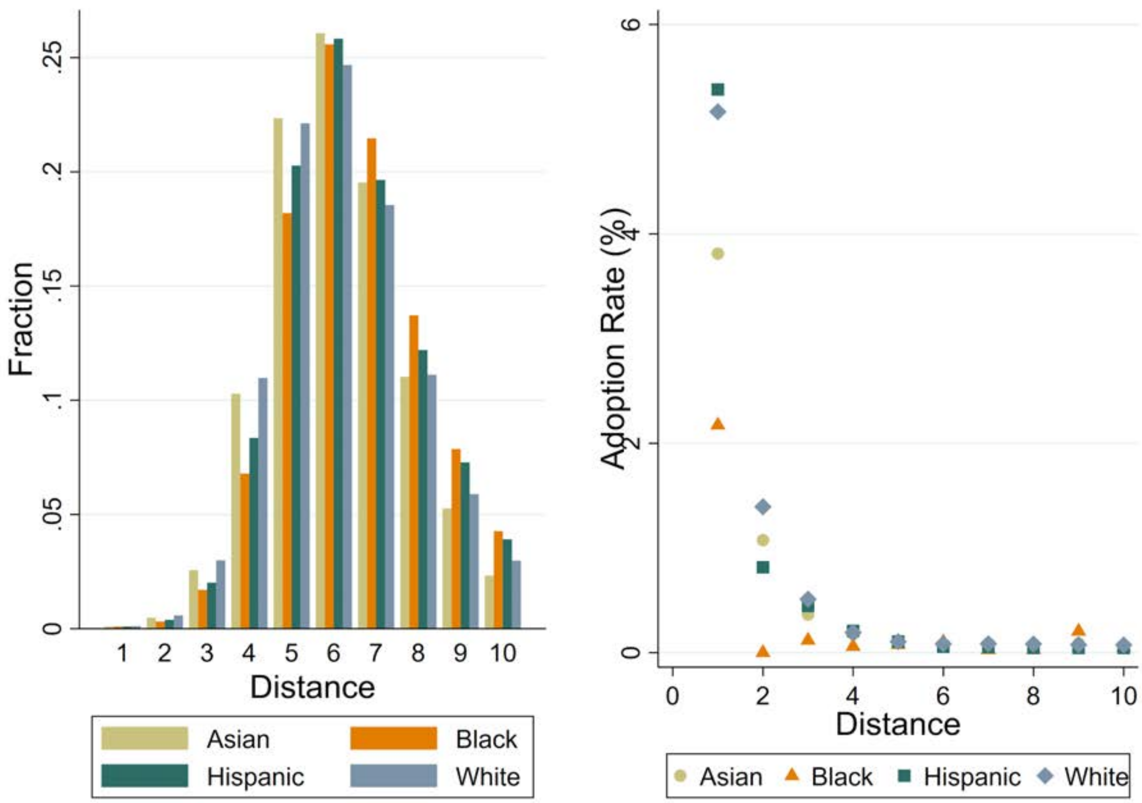
Notes: The left panel shows the adoption rates for female and male potential adopters at each network distance. The right panel considers adoption rates by different gender combinations of the potential adopter and his or her closest innovator. *FA* and *MA* represent a female or male potential adopter. *FI* and *MI* indicate that the closest innovator is female or male respectively. We move the points at distance 1 down by two units for readability.

Given that women are less likely to adopt new ideas, the gender composition of potential adopters at each distance to innovators is potentially an important determinant of adoption. Let *FI* and *MI* represent whether, for a given potential adopter, the closest innovator of a given idea is female or male. *FA* and *MA* indicate the gender of a potential adopter, yielding four combinations: *FI-FA*, *FI-MA*, *MI-FA* and *MI-MA*. The right panel of Figure 3 (which has a spliced Y-axis for readability), depicts the adoption rates for each of the four combinations by network distance. Two points are noteworthy. First, for both genders, people are more likely to adopt the ideas of network-proximate innovators of their own gender. Second, while at longer distances men are more likely to adopt the ideas of male innovators than women are to adopt the ideas of female innovators, at short distances (1 and 2), women are as likely to adopt the closet female innovator's ideas as men are to adopt the ideas of male innovators. Thus, it is not the case that women are uniformly less likely to adopt new ideas of women than men. Interestingly, at short distances, the lowest adoption rate is by women potential adopters of the ideas of male innovators, followed by men's adoption of women's ideas. Overall, men's lower adoption of women's ideas – and there are more male than female biomedical authors – explains the lower adoption of women's ideas, and women's reluctance to adopt men's ideas contributes to women's lower adoption of new, important ideas.

3.2 Descriptive results on race/ethnicity and experience

While our primary analysis is by gender, our methods apply to race and ethnicity and to career stage. Figure 4 provides estimates by race and ethnicity. Given that Asians, Blacks, and Hispanics are underrepresented, we classify an idea as having a substantial minority (Asian/Black/Hispanic) component if at least a fifth of innovators belong to that group. (The average idea is introduced by 5 innovators (median=4), so for the average idea, we are effectively assuming that 1 of the innovators is a minority.) Ideas that do not make that threshold are classified as non-Hispanic, White. The left panel of Figure 4 shows the distribution of potential adopters' network distance for the four groups of ideas. Ideas with Asian and White innovators are similar in terms of network distance; however, ideas with at least 20% Hispanic or Black innovators are farther away from potential adopters. In terms of adoption rates, the right panel shows that, beyond a distance of 1, the ideas of White inventor teams have the highest adoption rates followed by ideas from Asian, Hispanic, and then Black innovators. (Interestingly, at distance 1, ideas with Hispanic innovators are slightly more likely to be adopted than ideas by White innovators, while Asian and Black innovators' ideas are much less likely to be adopted.)

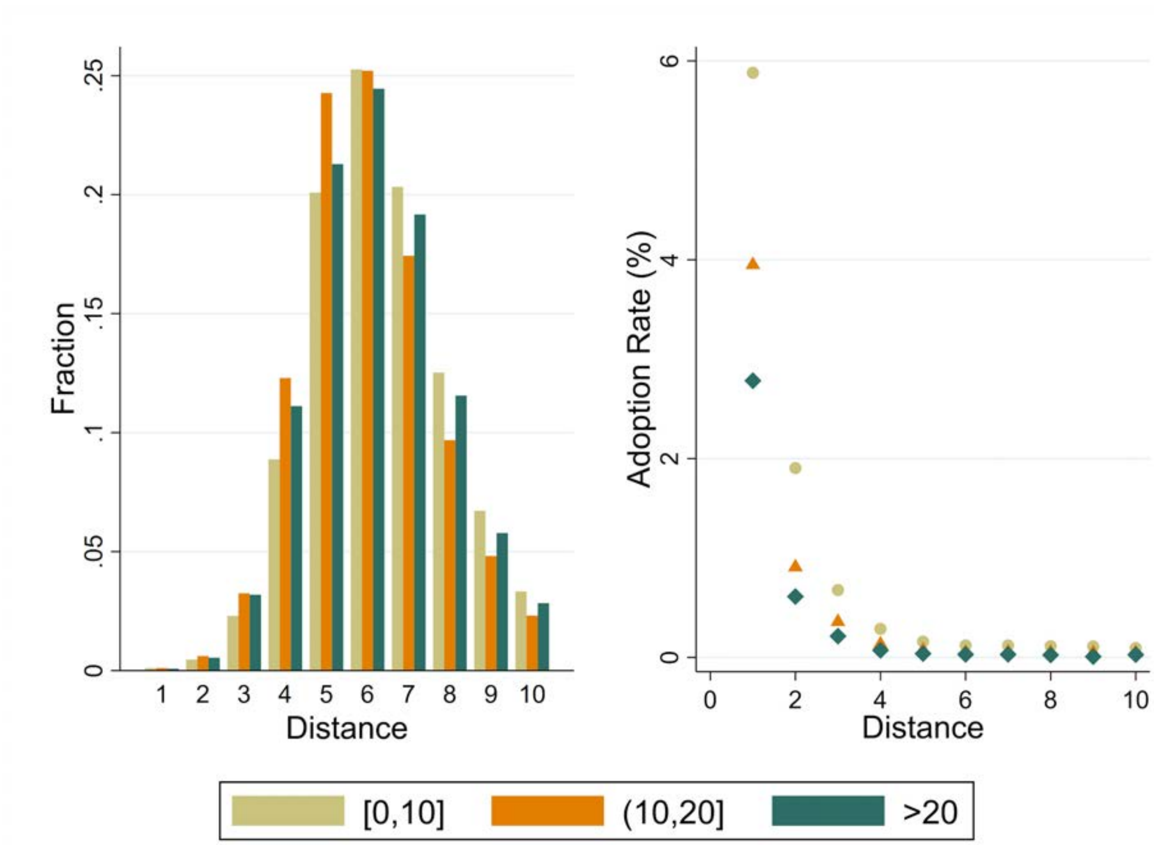
Figure 4. Race/ethnicity of the innovator



Notes: The left panel shows the distribution of network distance of potential adopters to ideas belonging to different racial / ethnic groups. If at least 20% of the innovators of an idea are Hispanic, we categorize this ideas as having a substantial Hispanic contribution. If less than 20% of the innovators are Hispanic, but 20% or more of the innovators are Black, we categorize this idea as having a substantial Black contribution. If neither of these conditions are met, and at least 20% of the innovators are Asian, then the idea is identified as having a substantial Asian contribution. The remaining ideas are labeled White. The right panel shows the adoption rate of ideas for these four groups by potential adopters at each network distance.

Early career innovators are frequently viewed as highly innovative, but more marginal within the research community. We conclude our descriptive analysis with an analysis of the role of experience in the diffusion of new ideas. We measure a researcher's experience using the number of years since his/her first publication. Innovators are younger than potential adopters on average - the average innovator has 9 years of experience compared to 16 years of experience for the average potential adopter. We divide ideas by average innovator experience into three groups: zero to 10 years, 11 to 20 years, and more than 20 years. The left panel of Figure 5 shows that potential adopters are closest to ideas from mid-career inventor teams (those with an average experience of 11-20 years). Strikingly, the right panel shows that ideas by the youngest innovator teams are most likely to be adopted at each network distance followed by ideas from mid-career and then older innovator teams. This is to say, mid-career innovators have better network connections, but ideas from the group of youngest innovators are more adopted for any given distance. The relative strength of the two competing forces determines which ideas have better overall adoption rates. We find that, overall, new ideas by the youngest have a higher adoption rate compared to those by mid-career innovators (0.17% vs 0.09%).

Figure 5. Career age of innovators



Notes: The left panel shows the distribution of network distance of potential adopters to the innovators of ideas based on the average experience of the innovators. The right panel shows the adoption rate of ideas for these three groups by potential adopters at each network distance.

3.3 Regression results

We investigate how innovator characteristics, potential adopter characteristics, and network distance interact to determine idea adoption rates using regressions of the form.

$$A_{i,j} = \alpha_0 + \sum_{d=1}^{5} \alpha_d \cdot 1(Distance_{i,j} = d) + \sum_{d=1}^{5} \beta_d \cdot 1(Distance_{i,j} = d) \cdot X_{\mathrm{j}}^{PA}$$
$$+ \sum_{d=1}^{5} \eta_d \cdot 1(Disance_{i,j} = d) \cdot X_{i,j}^{\ I} + \sum_{d=1}^{5} \lambda_d \cdot 1(Distance_{i,j} = d) \cdot X_j^{\ PA} \cdot X_{i,j}^{\ I} + \gamma_i + \rho_j +$$
$$u_{i,j} \tag{1}$$

where $A_{i,j}$ represents whether potential adopter j adopts (or independently adopts) idea i within five[9] years of its first introduction and $Distance_{i,j}$ represents researcher j's network distance to the closest innovator on idea i in its birth year. We include a full set of distance dummies to account for a flexible pattern in the relationship between network distance and adoption rates. Distances greater than 5 have similar coefficients, and have been combined into one group, which is the excluded / reference group. The term $X_j^{\ \ PA}$ represents potential adopter j's characteristics such as gender and experience[10]. The term $X_{i,j}^{\ I}$ represents the characteristics of the innovator on idea i who is closest to potential adopter j.[11] As discussed, our models contain idea fixed effects, represented by γ_i, which accounts for, among other factors, idea birth year differences, and potential adopter fixed effects, represented by ρ_j.

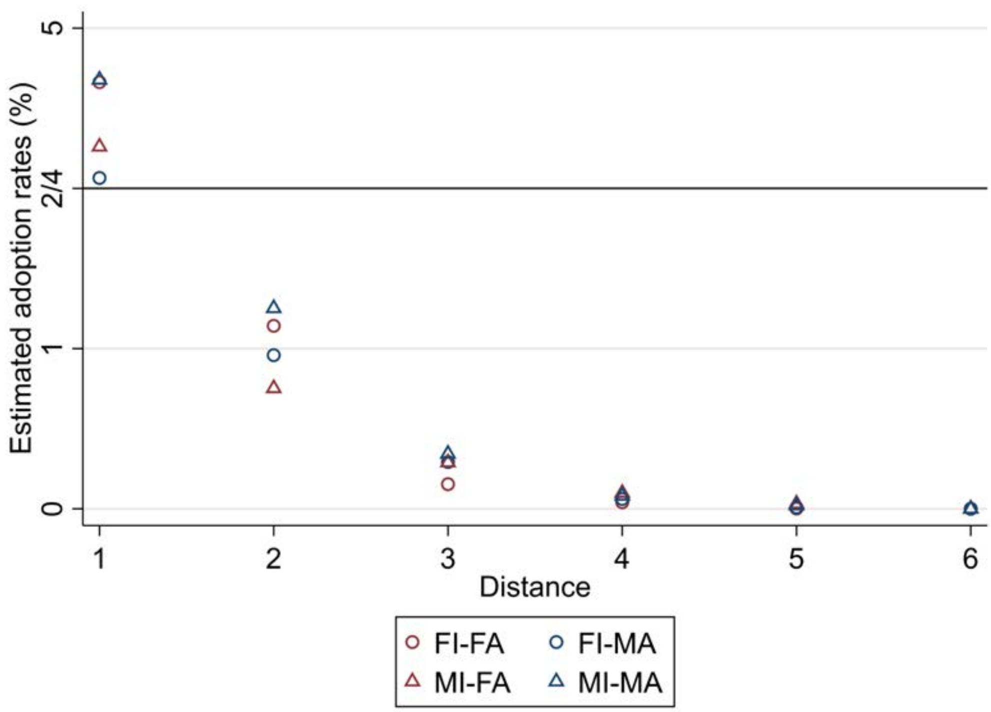
[9] Results using ten years instead of five years after introduction are similar.
[10] We do not interact distance with race and ethnicity because the sample sizes of Hispanics and Blacks are small and are even smaller when divided by network distance so that the estimates are imprecise.
[11] Averages are taken in case that there are multiple innovators who are the same distance from a given potential adopter.

There are three parameters of interest. First, α_d captures how adoption declines with network distance. Given the inclusion of idea fixed effects, β_d is identified from differences across people who are at different initial distances from the innovators of ideas. Ultimately, β_d likely captures the causal effect of distance as well as the decline in the relevance of work as proximity to the innovators of that idea increase. Second, η_d represents how the characteristics of the closest innovator interact with network distance. Given the inclusion of distance and idea fixed effects, η_d is identified from how the adoption probability for the same idea declines with distance for potential adopters who are closest to a male innovator compared to those who are closest to a female innovator. Lastly, λ_d describes how the interaction between innovator characteristics and potential adopter characteristics varies with network distance. Given the inclusion of potential adopter fixed effects (as well as distance and idea fixed effects), if X represents gender, λ_d is identified from the decline with distance in the probability of adopting an idea for male and female potential adopters when the closest innovator is a woman versus a man for the same adopter.

Table 2 provides estimates for our baseline regressions without interacted distance variables. The outcome variable in columns (1)-(2) is a binary indicator of adoption and that in columns (3)-(4) is independent adoption. We gradually expand the specification to incorporate additional control variables. All estimates are multiplied by 100 for scaling, so the coefficients can be interpreted as percentage point changes.

There are four main findings. First, the adoption probability decreases monotonically and substantially with network distance. The large magnitude of the coefficient for distance 1 indicates that immediate neighbors of the innovators of a new idea are substantially more likely to adopt the idea than those farther away. Including both sets of fixed effects does little to the coefficients on the distance dummies. Columns (4)-(6) present estimates for independent adoption. The baseline rate of independent adoption is lower than that for any adoption, and the effects of distance in Columns (4)-(6) tend to be smaller in magnitude than those for any adoption, but otherwise the two sets of estimates are similar.

Figure 6. Regression Estimates of Adoption by Gender of Innovator and Adopter



Notes: The figure plots regression-based adoption rates from Column 2 of Table 3 by different gender combinations of the potential adopter and his or her closest innovator. *FA* and *MA* represent a female or male potential adopter. *FI* and *MI* represent that the closest innovator is female or male respectively. We move the points at distance 1 down by two units for readability.

Second, potential adopters closest to an innovator who is female, Asian, or Hispanic, are less likely to adopt that idea. But the estimates become insignificant after we control for both individual fixed effects and idea fixed effects. Third, in terms of characteristics of potential adopters, female scientists are 0.035% less likely to adopt new ideas in the first 5 years. Considering that the overall adoption rate is 0.12%, this gender effect is quite sizable. Compared to the reference group of Whites, Blacks and Hispanics are less likely to adopt new ideas. After controlling for individual fixed effects (in Columns 2 and 5), gender, race and ethnicity are absorbed. Lastly, experience plays a different role for innovators and potential adopters. Ideas originated by younger or senior innovators are adopted more than those from middle-career innovators. Yet adoption rates decline monotonically with a potential adopter's experience, suggesting that young researchers are more likely to try out new ideas.

Table 3 shows how the effect of network distance declines by the gender of the innovator and of the potential adopter. The estimates on the distance dummies are qualitatively similar to those in Table 2. The negative coefficients on the interaction terms between network distance and the closest innovator being a woman suggest that being close to a female innovator significantly decreases a researcher's probability of adopting a new idea. Similarly, the interactions between network distance and potential adopter's gender indicate that even when women are proximate to innovators of a new idea, they are less likely to adopt the idea, which is consistent with Figure 3.

Interestingly, the model with three-way interactions between network distance, closest innovator gender, and potential adopter gender (reported in Column 2 and plotted in Figure 6) shows that, holding all else constant and compared to men, women are about as likely to adopt new ideas from female innovators as men are to adopt ideas from male innovators at short distances (1 and 2). However, the situation is different at larger distances, where both women and men (women more even so than men), are less likely to adopt a new idea when closest to a female innovator. In other words, women's ideas are even more restricted to nearby potential adopters. These findings suggest that low adoption rates for those closest to women innovators arise because men are less likely to adopt women's ideas even though they are only a step or two away from the innovators.

3.4 Decomposition analysis

To gauge the relative importance of factors in the lower adoption rates of women's new ideas, we conduct a series of decomposition exercises. Our first decomposition focuses on the adoption of ideas based on the gender of innovators. The second set of decompositions studies the gender of potential innovators and potential adopters. The third decomposition focuses on the gender mix of the entire team of innovators.

Overall, the adoption rate of FIs is 10.6% lower than that of MIs. We begin with the estimates in Column (1) of Table 3 and explore the effects of differences in the distribution of network distances between female innovators and male innovators, ignoring the genders of potential adopters, which are excluded from Column (1) of Table 3. As shown in Column 1 of Table 4, if we give FIs the network connections of MIs[12], while keeping the adoption rates unchanged at each distance, we can reduce the gender gap by 39% of the total difference. Women have fewer potential adopters at close distances and hence more at long distances (>5). Thus, giving FIs the network connections of MIs increases the adoption of FIs' ideas at short distances, but actually decreases adoption at greater (>5) distances. (The effects increase in distance even though adoption rates decline with distance because the number of potential adopters increases substantially with distance.) Overall, 61% of the gap in adoption between FIs had MIs is due to a lower adoption rate for the ideas of FIs compared to MIs at each distance.

Next, we continue to consider models of adoption that also consider the gender of potential adopters. There are two natural counterfactuals to consider – we can (A) give women the same

[12] We obtain the distribution of network distance for male innovators by restricting to new ideas with no female innovators.

gender mix of potential adopters as men or (B) give women the same mix of same-gender adopters as men. Our estimates for both decompositions use the estimates in Column (2) of Table 3, which includes potential adopter gender. If we give FIs the gender mix of potential adopters for MIs at each distance (i.e. more male researchers at each distance) but keep unchanged the distribution of distance and the adoption rates for all four combinations FI-FA, MI-FA, FI-MA, MI-MA, we find that the adoption rate of FIs' ideas is largely unchanged. Intuitively, for FIs, replacing FAs with MAs reduces adoption at short distances (because nearby women have higher adoption rates for FIs than MAs), but it raises adoption at greater distances (because distant men actually have higher adoption rates for FIs than women, up until a distance of 5+). These two effects offset each other, leaving the adoption rates of FIs relative to MIs largely unchanged. Moreover, the effects are quite small because the difference in the gender distribution of potential adopters is not that great.

Because we find that the gender match between innovators and potential adopters plays an important role, we explore an alternative counterfactual using this model. Namely, we consider giving FIs as many *same-gender* potential adopters as MIs (i.e. giving FIs as many FAs as MIs have MAs while preserving the total number of potential adopters at each distance). Because most potential adopters are men, this counterfactual has a much larger effect on the gender mix of potential adopters than the previous counterfactual. At short distances (1 and 2) increasing the number of FAs for FIs to match the number of MAs for MIs increases adoption for FIs slightly. However, this effect is more than offset when we increase the number of FAs for FIs to match the number of MAs for MIs at larger distances. Intuitively, at short distances, FAs have much higher adoption rates than MAs for FIs, so increasing the number of FAs at short distances increases adoption for FIs. However, at long distances, FAs actually have lower adoption rates for FIs than MAs do. Consequently, at greater distances (>2), giving FIs as many FAs as MIs have MAs actually reduces adoption for FIs. The overall effect of giving FIs the same share of same-gender adopters as MIs is actually to reduce adoption for FAs because the number of potential adopters at greater distances is so much larger than those nearby.

Our main findings from these decompositions are that the network positions of women and lower overall adoption for female innovators account respectively for roughly 40% and 60% of the lower adoption rates for female innovators. However, the gender mix of potential adopters is at best a small contributor to this gap. Rather, the gender gap arises even conditional on the gender mix of potential adopters

Our third decomposition returns to the analysis of teams from the introduction. Specifically, we study the gap in adoption between female-majority and male-majority ideas. By construction, people are more likely to be close to an FI than an MI on female majority ideas, but the two are not identical. How much does the lower adoption rate of FIs compared with MIs translate into differences in adoption rates of female majority ideas and female majority ideas? To answer this question, we consider two hypothetical ideas that represent respectively an average female majority idea and an average male majority idea. Both ideas have 5 innovators, which is the average number of innovators associated with a new idea. The hypothetical male majority idea has 12.7% (0.63) women innovators on the team; the hypothetical female majority idea has 68.5% (3.4) women innovators on the team. By using the sample average distribution of network connections for MIs and FIs and the adoption rates at each distance given the gender of potential adopters at each distance, we find less adoption for FIs explains 37.6% of the difference in adoption rates for these two hypothetical ideas.

3.5 Robustness checks

One concern with our estimates is that when multiple teams introduce the same idea in the same year, it is possible that both teams developed the idea independently or that one team developed the idea first and the other adopted it quickly. To address this issue, Appendix Tables 1 and 2 provide estimates for ideas introduced by single teams. The estimates are broadly similar to our main estimates. To probe the robustness of our results to the 5-year window for adoption, Appendix Tables 3 and 4 report estimates for adoption within 10 years of the introduction of an idea.

4. Conclusion

This paper studies the diffusion of new ideas through networks and how diffusion is related to the age, gender, race and ethnicity of innovators and potential adopters. Network proximity to innovators of a new idea is associated with a higher rate of adoption. Using a novel, two-way fixed effects strategy, we explore two reasons why ideas introduced by female scientists are under-utilized: first, female innovators are not as well-connected; second, even given a short distance, people are less likely to adopt ideas introduced by women, and this effect is particularly true for male potential adopters.

Although our analysis has focused on gender, we find concerning patterns for non-Hispanic Blacks and Hispanics, where their ideas are less likely to be adopted compared with non-Hispanic Whites. Our analysis suggests that this gap is not entirely because that they are more disadvantaged in terms of network positions. Younger researchers are disproportionately more likely to build on new ideas. Moreover, new ideas introduced by younger teams tend to be adopted even though young teams of innovators having worse network connections than middle-aged or older teams.

Table 1 Summary Statistics

	N	*Mean*	*SD*
Panel A. Innovators			
Female	19,616	0.26	0.44
Male	19,616	0.59	0.49
Asian	19,616	0.15	0.36
Black	19,616	0.006	0.077
Hispanic	19,616	0.035	0.185
White	19,616	0.80	0.40
Career age	19,616	8.94	8.89
Panel B. Potential adopters			
Female	536,987	0.33	0.47
Male	536,987	0.52	0.50
Asian	536,987	0.126	0.33
Black	536,987	0.006	0.08
Hispanic	536,987	0.040	0.20
White	536,987	0.83	0.38
Panel C. Potential adopter-idea pairs			
Adoption	7,417,333	0.12%	3.47%
Independent adoption	7,417,333	0.11%	3.33%
Network distance	7,417,333	6.15	0.85
Career age	7,417,333	16.26	11.05

Notes: Each observation is an innovator-idea in Panel A, a unique potential adopter in Panel B and a potential adopter-idea pair in Panel C. In Panel A, it is possible that one innovator may be associated with multiple new ideas. Adoption (independent adoption) is a binary measure of whether a researcher (independently) adopts a new idea. The mean and standard deviation of (independent) adoption are multiplied by 100 so that they are in percentage points.

Table 2 Baseline regressions

	Adoption			Independent adoption		
	(1)	(2)	(3)	(4)	(5)	(6)
Distance=1	4.69007***	4.66005***	4.53431***	3.08097***	3.11940***	3.00015***
	(0.2386)	(0.0419)	(0.0419)	(0.1958)	(0.0403)	(0.0403)
Distance=2	1.19889***	1.18386***	1.11433***	1.04288***	1.02971***	0.96673***
	(0.0562)	(0.0185)	(0.0186)	(0.0527)	(0.0178)	(0.0179)
Distance=3	0.38670***	0.35688***	0.32064***	0.35010***	0.32368***	0.29124***
	(0.0150)	(0.0084)	(0.0088)	(0.0143)	(0.0081)	(0.0084)
Distance=4	0.10943***	0.09010***	0.08167***	0.09860***	0.08090***	0.07369***
	(0.0051)	(0.0048)	(0.0052)	(0.0049)	(0.0046)	(0.0050)
Distance=5	0.02808***	0.01712***	0.02081***	0.02441***	0.01423***	0.01806***
	(0.0029)	(0.0035)	(0.0038)	(0.0028)	(0.0034)	(0.0037)
Characteristics of the Closest Innovators						
Female	-0.01515***	-0.00847*	-0.00150	-0.01479***	-0.00867**	-0.00124
	(0.0037)	(0.0043)	(0.0046)	(0.0036)	(0.0042)	(0.0044)
Asian	-0.04232***	-0.04475***	0.00396	-0.03753***	-0.03999***	0.00576
	(0.0058)	(0.0065)	(0.0069)	(0.0056)	(0.0062)	(0.0067)
Black	-0.01340	-0.02135	0.00414	-0.01651	-0.02118	0.00205
	(0.0309)	(0.0327)	(0.0335)	(0.0292)	(0.0315)	(0.0322)
Hispanic	-0.02631***	-0.02043**	-0.00016	-0.02563***	-0.02263**	-0.00675
	(0.0079)	(0.0093)	(0.0099)	(0.0075)	(0.0090)	(0.0095)
Career age	-0.00965***	-0.00802***	-0.00137**	-0.00881***	-0.00732***	-0.00118**
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)
Career age squared	0.00019***	0.00017***	0.00003**	0.00017***	0.00015***	0.00003*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Characteristics of potential adopters						
Female	-0.03528***			-0.03225***		
	(0.0029)			(0.0027)		
Asian	-0.00449			-0.00537		
	(0.0044)			(0.0042)		
Black	-0.02508*			-0.01759		
	(0.0151)			(0.0149)		
Hispanic	-0.01226*			-0.01219*		
	(0.0073)			(0.0070)		
Career age	-0.00118***			-0.00079**		
	(0.0003)			(0.0003)		
Career age squared	-0.00004***			-0.00004***		
	(0.0000)			(0.0000)		

Indiv. f.e.	No	Yes	Yes	No	Yes	Yes
Term f.e.	No	No	Yes	No	No	Yes
N	7,417,333	7,312,046	7,312,013	7,417,333	7,312,046	7,312,013
R squared	0.003	0.044	0.050	0.002	0.041	0.047

Notes: The outcome variable for Columns (1)-(3) is a binary variable for adoption or not, and that for Columns (4)-(6) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 5 years. All estimated coefficients are multiplied by 100.

Table 3 Gender and network distance

	Adoption		Independent adoption	
	(1)	(2)	(3)	(4)
Distance=1	4.57336***	4.67588***	3.01586***	3.15696***
	(0.0448)	(0.0515)	(0.0430)	(0.0495)
Distance=2	1.12981***	1.25227***	0.98623***	1.10689***
	(0.0198)	(0.0228)	(0.0191)	(0.0219)
Distance=3	0.33077***	0.34412***	0.30200***	0.31564***
	(0.0093)	(0.0107)	(0.0090)	(0.0102)
Distance=4	0.08623***	0.08132***	0.07774***	0.07169***
	(0.0055)	(0.0062)	(0.0053)	(0.0060)
Distance=5	0.02352***	0.02017***	0.02035***	0.01676***
	(0.0040)	(0.0046)	(0.0039)	(0.0044)
*FI*Distance=1*	-0.32551**	-0.61048***	-0.13737	-0.61244***
	(0.1265)	(0.1511)	(0.1216)	(0.1453)
*FI*Distance=2*	-0.12240**	-0.29280***	-0.15566***	-0.31247***
	(0.0558)	(0.0649)	(0.0536)	(0.0624)
*FI*Distance=3*	-0.07346***	-0.05031*	-0.07917***	-0.06638**
	(0.0246)	(0.0286)	(0.0236)	(0.0275)
*FI*Distance=4*	-0.03014**	-0.02022	-0.02674**	-0.01657
	(0.0136)	(0.0159)	(0.0131)	(0.0153)
*FI*Distance=5*	-0.01623	-0.01114	-0.01341	-0.00799
	(0.0102)	(0.0119)	(0.0098)	(0.0115)
FI	0.01056*	0.00972	0.00990*	0.00907
	(0.0056)	(0.0065)	(0.0054)	(0.0062)
*FA*Distance=1*		-0.41938***		-0.57746***
		(0.1041)		(0.1001)
*FA*Distance=2*		-0.50040***		-0.49303***
		(0.0458)		(0.0440)
*FA*Distance=3*		-0.05336**		-0.05449***
		(0.0208)		(0.0200)
*FA*Distance=4*		0.01954*		0.02407**
		(0.0116)		(0.0112)
*FA*Distance=5*		0.01331		0.01431*
		(0.0086)		(0.0082)
*FA*FI*		0.00309		0.00310
		(0.0120)		(0.0115)
*FA*FI*Distance=1*		1.01559***		1.66803***

		(0.2770)		(0.2663)
*FA*FI*Distance=2*		0.68271***		0.63063***
		(0.1270)		(0.1221)
*FA*FI*Distance=3*		-0.08420		-0.04524
		(0.0556)		(0.0534)
*FA*FI*Distance=4*		-0.03815		-0.03935
		(0.0303)		(0.0292)
*FA*FI*Distance=5*		-0.01948		-0.02072
		(0.0227)		(0.0218)
N	7,312,013	7,312,013	7,312,013	7,312,013
R squared	0.050	0.050	0.047	0.047

Notes: The outcome variable for Columns (1)-(2) is a binary variable for adoption or not, and that for Columns (3)-(4) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 5 years. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

Table 4. Decompositions of Differences in Adoption

Distance	1) Female innovators have the same distribution of network distance as male innovators	2) Counterfactual scenarios for the gender of potential adopters, by distance	
		2A) Same gender mix as MI	2B) Same gender-match as MI
1	9.1%	-0.37%	1.86%
2	9.9%	-0.15%	3.37%
3	14.9%	0.58%	-12.92%
4	21.7%	0.29%	-5.69%
5	29.6%	0.16%	-2.80%
>5	-46.7%	-0.35%	5.47%
Overall	38.49%	0.15%	-10.72%

Notes: Decomposition exercises based on estimates from Table 3. Column 1 is based on Column 1 of Table 3. Columns 2A and 2B are based on column 2 of Table 3. The adoption rate of FIs is 10.6% lower than that of MIs. The estimates show how much each change would reduce this gap as a share of the gap at the indicated distance and overall.

Appendix

The following tables repeat the regressions using the single-team subsample.

Appendix Table 1. Baseline regressions, single team ideas

	Adoption			Independent adoption		
	(1)	(2)	(3)	(4)	(5)	(6)
Distance=1	3.32986***	3.59127***	3.54726***	2.23005***	2.46538***	2.41613***
	(0.3527)	(0.0599)	(0.0601)	(0.2912)	(0.0577)	(0.0578)
Distance=2	0.67176***	0.66524***	0.67811***	0.59072***	0.57731***	0.59000***
	(0.0777)	(0.0273)	(0.0275)	(0.0731)	(0.0263)	(0.0264)
Distance=3	0.26722***	0.23788***	0.26015***	0.24478***	0.21679***	0.23847***
	(0.0224)	(0.0122)	(0.0126)	(0.0215)	(0.0117)	(0.0121)
Distance=4	0.08590***	0.06235***	0.08675***	0.07970***	0.05710***	0.08054***
	(0.0074)	(0.0066)	(0.0071)	(0.0072)	(0.0063)	(0.0069)
Distance=5	0.01936***	0.00273	0.02050***	0.01780***	0.00215	0.01910***
	(0.0037)	(0.0046)	(0.0050)	(0.0036)	(0.0044)	(0.0048)
Characteristics of the Closest Innovators						
Female	-0.00167	0.00112	-0.01064*	-0.00285	-0.00041	-0.01067*
	(0.0046)	(0.0054)	(0.0057)	(0.0044)	(0.0052)	(0.0055)
Asian	-0.02137***	-0.02141***	0.00443	-0.01857***	-0.01884**	0.00520
	(0.0067)	(0.0080)	(0.0086)	(0.0065)	(0.0077)	(0.0083)
Black	-0.03710	-0.04819	-0.03840	-0.03087	-0.03511	-0.02751
	(0.0260)	(0.0379)	(0.0394)	(0.0259)	(0.0365)	(0.0379)
Hispanic	0.02611**	0.03511***	0.01458	0.01921*	0.02613**	0.00483
	(0.0119)	(0.0115)	(0.0122)	(0.0112)	(0.0111)	(0.0118)
Career age	-0.00479***	-0.00444***	-0.00108	-0.00440***	-0.00413***	-0.00103
	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0006)	(0.0007)
Career age squared	0.00010***	0.00010***	0.00003*	0.00009***	0.00009***	0.00003*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Characteristics of potential adopters						
Female	-0.02356***			-0.02234***		
	(0.0034)			(0.0032)		
Asian	-0.00678			-0.00957**		
	(0.0050)			(0.0048)		
Black	-0.02396			-0.01881		
	(0.0162)			(0.0162)		
Hispanic	0.00616			0.00591		

	(0.0094)			(0.0091)		
Career age	-0.00024			-0.00027		
	(0.0004)			(0.0004)		
Career age squared	-0.00003***			-0.00003***		
	(0.0000)			(0.0000)		
Indiv. f.e.	No	Yes	Yes	No	Yes	Yes
Term f.e.	No	No	Yes	No	No	Yes
N	3,152,579	3,033,040	3,033,014	3,152,579	3,033,040	3,033,014
R squared	0.002	0.073	0.076	0.001	0.071	0.074

Notes: The outcome variable in Columns (1)-(3) is a binary variable on adoption or not, and that in Columns (4)-(6) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 5 years. All estimated coefficients are multiplied by 100. The sample is restricted to ideas that were introduced by a single team in their year of introduction.

Appendix Table 2. Gender and network distance, single team ideas

	Adoption		Independent adoption	
	(1)	(2)	(3)	(4)
Distance=1	3.52532***	3.68176***	2.48066***	2.44739***
	(0.0642)	(0.0741)	(0.0618)	(0.0713)
Distance=2	0.69027***	0.76546***	0.61920***	0.71487***
	(0.0292)	(0.0333)	(0.0281)	(0.0320)
Distance=3	0.26371***	0.27506***	0.24204***	0.25643***
	(0.0133)	(0.0151)	(0.0128)	(0.0145)
Distance=4	0.08857***	0.09369***	0.08213***	0.08601***
	(0.0075)	(0.0084)	(0.0072)	(0.0081)
Distance=5	0.02365***	0.02578***	0.02181***	0.02444***
	(0.0052)	(0.0059)	(0.0050)	(0.0057)
FI	-0.00330	-0.00363	-0.00294	-0.00255
	(0.0067)	(0.0077)	(0.0064)	(0.0074)
*FI*Distance=1*	0.16604	-0.58539***	-0.52749***	-1.57705***
	(0.1815)	(0.2203)	(0.1747)	(0.2120)
*FI*Distance=2*	-0.09978	-0.03172	-0.24352***	-0.27501***
	(0.0838)	(0.0969)	(0.0807)	(0.0933)
*FI*Distance=3*	-0.02636	-0.01366	-0.02703	-0.03166
	(0.0369)	(0.0428)	(0.0355)	(0.0412)
*FI*Distance=4*	-0.01139	0.00476	-0.00977	0.00543
	(0.0192)	(0.0222)	(0.0185)	(0.0214)
*FI*Distance=5*	-0.02279*	-0.01456	-0.01937	-0.01306
	(0.0134)	(0.0156)	(0.0129)	(0.0150)
*FA*Distance=1*		-0.62784***		0.13540
		(0.1484)		(0.1429)
*FA*Distance=2*		-0.32670***		-0.41629***
		(0.0691)		(0.0665)
*FA*Distance=3*		-0.04796		-0.06097**
		(0.0303)		(0.0292)
*FA*Distance=4*		-0.02134		-0.01657
		(0.0162)		(0.0156)
*FA*Distance=5*		-0.00881		-0.01102
		(0.0113)		(0.0109)
*FA*FI*		0.00085		-0.00181
		(0.0143)		(0.0138)
*FA*FI*Distance=1*		2.43089***		3.15004***

		(0.3904)		(0.3758)
*FA*FI*Distance=2*		-0.23771		0.16181
		(0.1931)		(0.1859)
*FA*FI*Distance=3*		-0.04627		0.02082
		(0.0840)		(0.0809)
*FA*FI*Distance=4*		-0.06224		-0.05895
		(0.0435)		(0.0419)
*FA*FI*Distance=5*		-0.03100		-0.02356
		(0.0302)		(0.0290)
N	3,033,014	3,033,014	3,033,014	3,033,014
R squared	0.076	0.076	0.074	0.074

Notes: The outcome variable in Columns (1)-(2) is a binary variable on adoption or not, and that in Columns (3)-(4) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 5 years. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects. The sample is restricted to ideas that were introduced by a single team in their year of introduction.

The following tables provide robustness checks using the outcome variable for adoption within 10 years following the introduction of new ideas.

Appendix Table 3. Baseline regressions (adoption within 10 years)

	Adoption			Independent adoption		
	(1)	(2)	(3)	(4)	(5)	(6)
Distance=1	6.47232***	6.56447***	6.38140***	4.33168***	4.43425***	4.25607***
	(0.2786)	(0.0637)	(0.0635)	(0.2322)	(0.0620)	(0.0619)
Distance=2	2.25819***	2.18658***	2.07066***	1.99516***	1.93206***	1.82402***
	(0.0781)	(0.0284)	(0.0285)	(0.0738)	(0.0276)	(0.0277)
Distance=3	0.74240***	0.65922***	0.61870***	0.67850***	0.59834***	0.56142***
	(0.0213)	(0.0129)	(0.0134)	(0.0205)	(0.0125)	(0.0130)
Distance=4	0.22104***	0.16413***	0.17732***	0.20854***	0.15287***	0.16664***
	(0.0076)	(0.0073)	(0.0080)	(0.0074)	(0.0071)	(0.0078)
Distance=5	0.05295***	0.01569***	0.04208***	0.04887***	0.01289**	0.03896***
	(0.0045)	(0.0054)	(0.0058)	(0.0044)	(0.0052)	(0.0057)
Characteristics of the Closest Innovators						
Female	-0.03883***	-0.03011***	-0.00941	-0.03783***	-0.02930***	-0.00997
	(0.0057)	(0.0066)	(0.0069)	(0.0055)	(0.0064)	(0.0068)
Asian	-0.11076***	-0.13698***	-0.02917***	-0.10609***	-0.13205***	-0.02932***
	(0.0087)	(0.0099)	(0.0106)	(0.0084)	(0.0096)	(0.0103)
Black	0.01766	0.02571	0.03498	0.02792	0.03547	0.04237
	(0.0512)	(0.0500)	(0.0511)	(0.0507)	(0.0487)	(0.0498)
Hispanic	-0.03507***	-0.03485**	-0.00082	-0.03290***	-0.03255**	-0.00279
	(0.0125)	(0.0143)	(0.0151)	(0.0122)	(0.0139)	(0.0147)
Career age	-0.02189***	-0.01989***	-0.00469***	-0.02108***	-0.01935***	-0.00497***
	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.0008)	(0.0009)
Career age squared	0.00043***	0.00042***	0.00011***	0.00041***	0.00040***	0.00011***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Characteristics of potential adopters						
Female	-0.07183***			-0.06770***		
	(0.0044)			(0.0043)		
Asian	-0.01699**			-0.01658***		
	(0.0066)			(0.0064)		
Black	-0.04235*			-0.02917		
	(0.0239)			(0.0238)		
Hispanic	-0.03240***			-0.03172***		
	(0.0110)			(0.0107)		

Career age	0.00183***			0.00199***		
	(0.0005)			(0.0005)		
Career age squared	-0.00017***			-0.00017***		
	(0.0000)			(0.0000)		
Indiv. f.e.	No	Yes	Yes	No	Yes	Yes
Term f.e.	No	No	Yes	No	No	Yes
N	7,417,387	7,312,328	7,312,294	7,417,387	7,312,328	7,312,294
R squared	0.003	0.043	0.054	0.002	0.041	0.052

Notes: The outcome variable in Columns (1)-(3) is a binary variable on adoption or not, and that in Columns (4)-(6) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 10 years. All estimated coefficients are multiplied by 100.

Appendix Table 4. Gender and network distance (adoption within 10 years)

	Adoption			
	(1)	（2）	（3）	(4)
Distance=1	6.54984***	6.77339***	4.38007***	4.56841***
	(0.0679)	(0.0782)	(0.0661)	(0.0762)
Distance=2	2.09161***	2.17466***	1.84556***	1.92167***
	(0.0304)	(0.0348)	(0.0296)	(0.0339)
Distance=3	0.62850***	0.66606***	0.57042***	0.60807***
	(0.0142)	(0.0162)	(0.0138)	(0.0158)
Distance=4	0.18126***	0.18830***	0.17054***	0.17804***
	(0.0084)	(0.0095)	(0.0082)	(0.0092)
Distance=5	0.04296***	0.03815***	0.03954***	0.03495***
	(0.0061)	(0.0070)	(0.0060)	(0.0068)
*FI*Distance=1*	-1.36499***	-1.65844***	-1.00936***	-1.27984***
	(0.1913)	(0.2286)	(0.1863)	(0.2227)
*FI*Distance=2*	-0.15936*	-0.00276	-0.16445**	-0.03512
	(0.0846)	(0.0976)	(0.0824)	(0.0950)
*FI*Distance=3*	-0.06436*	-0.04312	-0.05869	-0.03276
	(0.0374)	(0.0437)	(0.0364)	(0.0425)
*FI*Distance=4*	-0.01760	-0.04027*	-0.01796	-0.04433*
	(0.0206)	(0.0242)	(0.0201)	(0.0235)
*FI*Distance=5*	0.00451	0.02497	0.00642	0.02681
	(0.0155)	(0.0181)	(0.0151)	(0.0177)
FI	-0.00154	-0.00143	-0.00284	-0.00291
	(0.0085)	(0.0098)	(0.0083)	(0.0096)
*FA*Distance=1*		-0.90788***		-0.76478***
		(0.1575)		(0.1534)
*FA*Distance=2*		-0.34305***		-0.31437***
		(0.0704)		(0.0686)
*FA*Distance=3*		-0.15175***		-0.15220***
		(0.0317)		(0.0309)
*FA*Distance=4*		-0.02752		-0.02939*
		(0.0177)		(0.0172)
*FA*Distance=5*		0.01890		0.01799
		(0.0130)		(0.0127)
*FA*FI*		-0.00086		-0.00018
		(0.0183)		(0.0178)
*FA*FI*Distance=1*		1.12978***		1.02682**

		(0.4184)		(0.4076)
*FA*FI*Distance=2*		-0.61730***		-0.50882***
		(0.1955)		(0.1905)
*FA*FI*Distance=3*		-0.06688		-0.08398
		(0.0842)		(0.0821)
*FA*FI*Distance=4*		0.08324*		0.09685**
		(0.0461)		(0.0449)
*FA*FI*Distance=5*		-0.07570**		-0.07538**
		(0.0345)		(0.0336)
N	7,312,294	7,312,294	7,312,294	7,312,294
R squared	0.054	0.054	0.052	0.052

Notes: The outcome variable in Columns (1)-(2) is a binary variable on adoption or not, and that in Columns (3)-(4) is a binary variable for independent adoption or not. Independent adoption means that a potential adopter uses the idea in an article that is not coauthored with any of the original innovators of the idea. Adoption is measured by whether a person adopts each new idea within 10 years. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

References

1. Agrawal, Ajay, Iain Cockburn and John McHale. Gone but Not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships. Journal of Economic Geography, 2006, 6 (5): 571-591.

2. Alcácer, Juan and Michelle Gittelman. Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations. The Review of Economics and Statistics, 2006, 88 (4): 774-779.

3. Almeida, Paul and Bruce Kogut. Localization of Knowledge and the Mobility of Engineers in Regional Networks. Management Science, 1999, 45 (7): 905-917.

4. Association of American Medical Colleges. The Diversity Research Forum: Getting to Institutional Excellence: Ensuring the Integration of Diversity in Academic Medicine [R]. 2009.

5. Azoulay, Pierre, Joshua Graff Zivin and Jialan Wang. Superstar Extinction. Quarterly Journal of Economics, 2010, 25: 549-589.

6. Benjamin, Gomes-Casseres, John Hagedoorn and Adam Jaffe. Do Alliances Promote Knowledge Flows? Journal of Financial Economics, 2006, 80 (1): 5-33.

7. Borjas, George J. and Kirk B. Doran. The Collapse of the Soviet Union and the Productivity of American Mathematicians. The Quarterly Journal of Economics, 2012, 127 (3): 1143-1203.

8. Borjas, George J and Kirk B Doran. Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors. Review of Economics and Statistics, 2014, (0).

9. Borjas, George J and Kirk B Doran. How High-Skill Immigration Affects Science: Evidence from the Collapse of the USSR. Innovation Policy and the Economy, 2015, 15 (1): 1-25.

10. Buffington, Catherine, Benjamin Cerf, Christina Jones, and Bruce A. Weinberg. STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census. American Economic Review, 2016. 106 (5):333-38

11. Card, David, Stefano DellaVigna, Patricia Funk and Nagore Iriberri. Are Referees and Editors in Economics Gender Neutral? The Quarterly Journal of Economics, 2020, 135 (1): 269-327.

12. Catalini, Christian. Microgeography and the Direction of Inventive Activity [R]. Rotman School of Management Working Paper No. 2126890, 2017.

13. Ceci, Stephen J, Donna K Ginther, Shulamit Kahn and Wendy M Williams. Women in Academic Science: A Changing Landscape. Psychological science in the public interest, 2014, 15 (3): 75-141.

14. Cikara, M., L. Rudman, and S. Fiske. Dearth by a Thousand Cuts? Accounting for Gender Differences in Top-Ranked Publication Rates in Social Psychology. Journal of Social Issues. 68(2): 263-285 (2012).

15. Collins, H. Gravity's Shadow. The Search for Gravitational Waves. 2004. Chicago: University of Chicago Press.

16. Fong, Eric A. and Allen W. Wilhite. Authorship and Citation Manipulation in Academic Research. PLOS ONE, 2017, 12 (12).

17. Freeman, R. B. and W. Huang. Collaboration: Strength in diversity. Nature, 2014, 513(7518): 305-305.

18. Gilbert, Nigel G. References as Persuasion. Social Studies of Science, 1977. 7(1): 113-22.

19. Ginther, Donna K and Shulamit Kahn. Does Science Promote Women? Evidence from Academia 1973-2001 [A]. In (eds.), Science and Engineering Careers in the United States: An Analysis of Markets and Employment[M]. University of Chicago Press, 2009: 163-194.

20. Ginther, Donna, K. Schaffer, Walter T., Schnell, Joshua, Masimore, Beth, Liu, Faye Haak, Laurel and Raynard Kington.2011. Race, Ethnicity, and NIH Research Awards. Science, August 19,333(6045): 1015-1019.

21. Ginther, Donna K., Shulamit Kahn and Walter T. Schaffer. Gender, Race, Ethnicity and NIH R01 Research Awards: Is There Evidence of a Double Bind? Academic Medicine, 2016, 91 (8): 1098-1107.

22. Glaeser, Edward L., Hedi D. Kallal, Jose A Scheinkman, and Andrei Shleifer. Growth in Cities. Journal of Political Economy, 1992, 100 (6): 1126-1152.

23. Glenn Ellison and Edward L. Glaeser. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. Journal of Political Economy, 1997, 105 (5): 889-927.

24. Halpern, Diane F, Camilla P Benbow, David C Geary, Ruben C Gur, Janet Shibley Hyde and Morton Ann Gernsbacher. The Science of Sex Differences in Science and Mathematics. Psychological science in the public interest, 2007, 8 (1): 1-51.

25. Jacobs, Jane. The Death and Life of Great American Cities [M]. New York: Vantage Books, 1961.

26. Jaffe, Adam B. Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value. The American Economic Review, 1986, 76 (5): 984-1001.

27. Jaffe, Adam B, Manuel Trajtenberg and Rebecca Henderson. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. the Quarterly journal of Economics, 1993, 108 (3): 577-598.

28. Jones, Benjamin F. Age and Great Invention. The Review of Economics and Statistics, 2010, 92 (1): 1-14.

29. Jones, Benjamin F. and Bruce A. Weinberg. Age Dynamics in Scientific Creativity. Proceedings of the National Academy of Sciences, 2011, 108 (47): 18910.

30. Kaiser, David. Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics [M]. Chicago: University of Chicago Press, 2005.

31. Krugman, Paul. Increasing Returns and Economic Geography. Journal of Political Economy, 1991, 99 (3): 483-499.

32. C. Lerchenmüller, M. J. Lerchenmueller, O. Sorenson "Long-term analysis of sex differences in prestigious authorships in cardiovascular research supported by the National Institutes of Health." Circulation, 2018, 137: 880-882.

33. Levin, Daniel Z. and Rob Cross. The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. Management Science, 2004, 50 (11): 1477-1490.

34. Koffi, Marelene. Innovative Ideas and Gender Inequality. Working Paper, 2021.

35. Mairesse, Jacques and Laure Turner. Measurement and Explanation of the Intensity of Co-Publication in Scientific Research: An Analysis at the Laboratory Level. National Bureau of Economic Research Working Paper Series, 2005, No. 11172.

36. Marschke, Gerald, Allison Nunez, Bruce A. Weinberg, and Huifeng Yu. Last Place? The Intersection of Ethnicity, Gender, and Race in Biomedical. AEA papers and proceedings. American Economic Association, 2018. 108 (5):222-227.

37. Mehta, Aditi, Marc Rysman and Tim Simcoe. Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion. Journal of Applied Econometrics, 2010, 25 (7): 1179-1204.

38. Merton, R.K. The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press, 1973.

39. Moed, Henk F. Citation Analysis in Research Evaluation Amsterdam: Springer Netherlands. 2005.

40. Moed, Henk F. and Eugene Garfield. In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter. Scientometrica, 2004, 60, No. 3 (2004) 295-303.

41. Moed, Henk F. and Thed van Leeuwen. “Impact Factors Can Mislead.” Nature, 1995, 381 (16 May): 186.

42. Oh, Sam S., Joshua Galanter, Neeta Thakur, Maria Pino-Yanes, Nicolas E. Barcelo, Marquitta J. White, Danielle M. de Bruin, Ruth M. Greenblatt, Kirsten Bibbins-Domingo, Alan H. B. Wu, Luisa N. Borrell, Chris Gunter, Neil R. Powe and Esteban G. Burchard. Diversity in Clinical and Biomedical Research: A Promise yet to Be Fulfilled. PLOS Medicine, 2015, 12 (12).

43. Packalen, Mikko and Jay Bhattacharya. Age and the Trying out of New Ideas. Journal of Human Capital, 2019, 13 (2): 341-373.

44. Peri, Giovanni. Determinants of Knowledge Flows and Their Effect on Innovation. Review of Economics and Statistics, 2005, 87 (2): 308-322.

45. Reagans, Ray and Ezra W Zuckerman. Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. Organization science, 2001, 12 (4): 502-517.

46. Rhoten, Diana, Erin O'Connor, and Edward J. Hackett. The Act of Collaborative Creation and the Art of Integrative Creativity: Originality, Disciplinarity and Interdisciplinarity. Thesis Eleven, 2009, 96 (1):83-108.

47. Romer, Paul M. Endogenous Technological Change. Journal of political Economy, 1990, 98 (5, Part 2): S71-S102.

48. Ross, Matthew B.; Britta Glennon; Raviv Murciano-Goroff; Enrico Berkes; Bruce A. Weinberg; and Julia I. Lane. Rosalind Franklin at Scale: Credit and Women in Science. Working Paper.

49. Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. Gender Differences in Recognition for Group Work, Journal of Political Economy, 2021, 129, No. 1: 101–147.

50. Singh, Jasjit. Collaborative Networks as Determinants of Knowledge Diffusion Patterns. Management Science, 2005, 51 (5): 756-770.

51. Staudt, Joseph, Huifeng Yu, Robert P. Light, Gerald Marschke, Katy Börner and Bruce A. Weinberg. High-Impact and Transformative Science (Hits) Metrics: Definition, Exemplification, and Comparison. PLOS ONE, 2018, 13 (7): e0200597.

52. Stephan, Paula E. and Sharon G. Levin. Age and the Nobel Prize Revisited. Scientometrics, 1993, 28 (3): 387-399.

53. Thompson, P. and M. Fox-Kean. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. American Economic Review, 2005, 95(1): 450-460.

54. Torvik, Vetle I. M. Weeber, Donald R. Swanson, Neil. R. Smalheiser. Journal of the Association for Information Science and Technology, 2005, 56 (2), 140-158.

55. Torvik, Vetle I and Neil R. Smalheiser. ACM Transactions on Knowledge Discovery from Data, ACM Transactions on Knowledge Discovery from Data (TKDD) 2009,3(3): 1-29.

56. Torvik Vetle I, Agarwal Sneha. Ethnea -- an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. International Symposium on Science of Science March 22-23, 2016. Library of Congress, Washington DC, USA.

57. Waldinger, Fabian. Quality Matters: The Expulsion of Professors and the Consequences for Phd Student Outcomes in Nazi Germany. Journal of Political Economy, 2010, 118 (4): 787-831.

58. Waldinger, Fabian. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. The Review of Economic Studies, 2012, 79 (2): 838-861.

59. Wennerås, C. and A. Wold. Nepotism and Sexism in Peer-Review. Nature, 1997, 7: 46-52.

60. West, Jevin D., J. Jacquet, M. M. King, S. J. Correll, C. T. Bergstrom, The role of gender in scholarly authorship. PLoS ONE 8(7).

61. Woolgar, S. Beyond the Citation Debate: Towards a Sociology of Measurement Technologies and Their Use in Science Policy. Science and Public Policy. 1991, 18: 319-26.

62. Wuchty, Stefan, Benjamin F Jones and Brian Uzzi. The Increasing Dominance of Teams in Production of Knowledge. Science, 2007, 316 (5827): 1036-1039.

63. Zacchia, Paolo. Knowledge Spillovers through Networks of Scientists. The Review of Economic Studies, 2019.

64. Zucker, Lynne G., Michael R. Darby and Marilynn B. Brewer. Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. The American Economic Review, 1998, 88 (1): 290-306.

65. Zuckerman, Harriett. Scientific Elite. [M]. New York: Free Press, 1977.